A Strong Limit Theorem for Two-Time-Scale Fucntional Stochastic Differential Equations

Jianhai Bao,* Qingshuo Song,† George Yin,‡ Chenggui Yuan§

August 31, 2015

Abstract

This paper focuses on a class of two-time-scale functional stochastic differential equations, where the phase space of the segment processes is infinite-dimensional. It develops ergodicity of the fast component and obtains a strong limit theorem for the averaging principle in the spirit of Khasminskii's averaging approach for the slow component.

Keywords: Two time scale, functional differential equation, exponential ergodicity, invariant measure, averaging principle

AMS Subject Classification: 60H15, 60J25, 60H30, 39B82

*Department of Mathematics, Central South University, Changsha, Hunan, 410083, P.R.China, jianhaibao13@gmail.com

†Department of Mathematics, City University of Hong Kong, Hong Kong, qingsong@cityu.edu.hk

‡Department of Mathematics, Wayne State University, Detroit, MI 48202, USA, gyin@math.wayne.edu

§Department of Mathematics, Swansea University, Singleton Park, SA2 8PP, UK, C.Yuan@swansea.ac.uk

1 Introduction

Having a wide range of applications in science and engineering (e.g., van Kampen [27]), singularly perturbed systems, have been investigated extensively recently; see, for instance, Freidlin-Wentzell [11], and Yin-Zhang [29]. Singularly perturbed systems usually exhibit multi-scale behavior owing to inherent rates of changes of the systems or different rates of interactions of subsystems and components. To reflect the slow and fast motions in the underlying systems, a time-scale separation parameter $\varepsilon \in (0,1)$ is often introduced. Due to the multi-scale property, it is frequently difficult to deal with such systems using a direct approach. As a result, it is foremost important to reduce their complexity. The averaging principle pioneered by Khasminskii [18] for a class of diffusions provides an effective way to reduce the complexity of the systems. For systems in which both fast and slow components co-exist, the idea of the averaging approach reveals that there is a limit dynamic system given by the average of the slow component with respect to the invariant measure of the fast component that is an ergodic process. The averaging equation approximates the slow component in a suitable sense whenever $\varepsilon \downarrow 0$ leading to a substantial reduction of computational complexity. The work [18] by Khasminskii inspired much of the subsequent development. To date, there have been a vast literature on the study of for multi-scale stochastic dynamic systems (see, e.g., the monograph [16]). For strong/weak convergence in averaging principle, we refer to, e.g., Givon et al. [13], Liu [22], Liu-Yin [23], and Yin-Zhang [29] for stochastic differential equations (SDEs), and Blömker et al. [4], Bréhier [5], Cerrai [6], Fu et al. [12], and Kuksin-Piatnitski [19] for stochastic partial differential equations (SPDEs); With regarding to numerical methods, we refer to, e.g., E et al. [10] and Givon et al. [14]; As for related control and filtering problems, we mention, e.g., Kushner [20, 21]. Concerning large deviations, we refer to, e.g., Kushner [20], and Veretennikov [28].

The aforemention references are all concerned with systems without "memory". Nevertheless, more often than not, dynamic systems with delay are un-avoidable in a wide variety of applications in science and engineering, where the dynamics are subject to propagation of delays. In response to the great needs, there is also an extensive literature on functional SDEs; see, e.g., the monographs [24, 25].

In contrast to the rapid progress in two-time-scale systems and differential delay equations, the study on averaging principles for functional SDEs is still in its infancy. Compared with the existing literature, for such systems, one of the outstanding issues is the phase space of the segment processes is infinite-dimensional, which makes the goal of obtaining a strong

limit theorem for the averaging principle a very difficult task. This work aims to take the challenges and to establish a strong limit theorem for the averaging principles for a range of two-time-scale functional SDEs.

The rest of the paper is organized as follows. Section 2 presents the setup of the problem we wish to study. The ergodicity of the frozen equation with memory is obtained in Section 3. Section 4 constructs some auxiliary two-time-scale stochastic systems with memory and provides a number of preliminary lemmas. Section 5 derives a strong limit theorem for the averaging principle in the spirit of Khasminskii's approach for the slow component.

Before proceeding further, a word of notation is in order. Throughout the paper, generic constants will be denoted by c; we use the shorthand notation $a \lesssim b$ to mean $a \le cb$, we use $a \lesssim_T b$ to emphasize the constant c depends on T.

2 Formulation

For integers $n, m \ge 1$, let $(\mathbb{R}^n, |\cdot|, \langle \cdot, \cdot \rangle)$ be an n-dimensional Euclidean space, and $\mathbb{R}^n \otimes \mathbb{R}^m$ denote the collection of all $n \times m$ matrices with real entries. For an $A \in \mathbb{R}^n \otimes \mathbb{R}^m$, $\|A\|$ stands for its Frobenius matrix norm. For an interval $I \subset (-\infty, \infty)$, $C(I; \mathbb{R}^n)$ means the family of all continuous functions from $I \mapsto \mathbb{R}^n$. For a fixed $\tau > 0$, let $\mathscr{C} = C([-\tau, 0]; \mathbb{R}^n)$, endowed with the uniform norm $\|\cdot\|_\infty$. For $h(\cdot) \in C([-\tau, \infty); \mathbb{R}^n)$ and $t \ge 0$, define the segment $h_t \in \mathscr{C}$ by $h_t(\theta) = h(t + \theta)$, $\theta \in [-\tau, 0]$.

Introducing a time-scale separation parameter $\varepsilon \in (0, 1)$, we consider two-time-scale systems of functional stochastic differential equations (SDEs) of the following form

$$\text{(2.1)} \qquad \mathrm{d}X^\varepsilon(t) = b_1(X_t^\varepsilon, Y_t^\varepsilon)\mathrm{d}t + \sigma_1(X_t^\varepsilon)\mathrm{d}W_1(t), \quad t > 0, \quad X_0^\varepsilon = \xi \in \mathscr{C},$$

and

$$\text{(2.2)} \quad \mathrm{d}Y^\varepsilon(t) = \frac{1}{\varepsilon} b_2(X_t^\varepsilon, Y^\varepsilon(t), Y^\varepsilon(t-\tau))\mathrm{d}t + \frac{1}{\sqrt{\varepsilon}} \sigma_2(X_t^\varepsilon, Y^\varepsilon(t), Y^\varepsilon(t-\tau))\mathrm{d}W_2(t), \ t > 0$$

with the initial value $Y_0^\varepsilon = \eta \in \mathscr{C}$, where $b_1 : \mathscr{C} \times \mathscr{C} \mapsto \mathbb{R}^n$, $b_2 : \mathscr{C} \times \mathbb{R}^n \times \mathbb{R}^n \mapsto \mathbb{R}^n$, $\sigma_1 : \mathscr{C} \mapsto \mathbb{R}^n \otimes \mathbb{R}^m$, $\sigma_2 : \mathscr{C} \times \mathbb{R}^n \times \mathbb{R}^n \mapsto \mathbb{R}^n \otimes \mathbb{R}^m$ are Gâteaux differentiable, $(W_1(t))_{t \ge 0}$ and $(W_2(t))_{t \ge 0}$ are two mutually independent m-dimensional Brownian motions defined on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$, equipped with $(\mathscr{F}_t)_{t \ge 0}$, a family of filtrations satisfying the usual conditions (i.e., for each $t \ge 0$, $\mathscr{F}_t = \mathscr{F}_{t+} := \bigcap_{s > t} \mathscr{F}_s$, and $\mathscr{F}_0$ contains all $\mathbb{P}$-null sets). As usual, for two-time-scale systems (2.1) and (2.2), $X^\varepsilon(t)$ is called the slow component, while $Y^\varepsilon(t)$ is called the fast component.

We denote by $\nabla^{(i)}$ the gradient operators for the i-th component. Throughout the paper, for any $\chi, \phi \in \mathscr{C}$ and $x, x', y, y' \in \mathbb{R}^n$, we assume that

(**A1**) $\nabla b_1 = (\nabla^{(1)} b_1, \nabla^{(2)} b_1)$ is bounded, and there exists an $L > 0$ such that

$$|b_1(\chi, \phi)| \le L(1 + \|\chi\|_\infty) \quad \text{and} \quad \|\sigma_1(\phi) - \sigma_1(\chi)\| \le L\|\phi - \chi\|_\infty.$$

(**A2**) $\nabla b_2 = (\nabla^{(1)} b_2, \nabla^{(2)} b_2, \nabla^{(3)} b_2)$ and $\nabla \sigma_2 = (\nabla^{(1)} \sigma_2, \nabla^{(2)} \sigma_2, \nabla^{(3)} \sigma_2)$ are bounded.

(**A3**) There exist $\lambda_1 > \lambda_2 > 0$, independent of χ, such that

$$\begin{aligned} &2\langle x - x', b_2(\chi, x, y) - b_2(\chi, x', y')\rangle + \|\sigma_2(\chi, x, y) - \sigma_2(\chi, x', y')\|^2 \\ &\quad \le -\lambda_1 |x - x'|^2 + \lambda_2 |y - y'|^2. \end{aligned}$$

(**A4**) For the initial value $X_0^\varepsilon = \xi \in \mathscr{C}$ of (2.1), there exists a $\lambda_3 > 0$ such that

$$|\xi(t) - \xi(s)| \le \lambda_3 |t - s|, \quad s, t \in [-\tau, 0].$$

Let us comment the assumptions (**A1**)-(**A4**) above. From (**A1**) and (**A2**), the gradient operators ∇b_1, ∇b_2, and $\nabla \sigma_2$ are bounded, respectively, so that b_1, b_2, and σ_2 are Lipschitz. Then, both (2.1) and (2.2) are well posed (see, e.g., [24, Theorem 2.2, P.150]). While, (**A3**) is imposed to analyze the ergodic property of the frozen equation (see Theorem 3.1 below), guarantee the Lipschitz property of $\overline{b}_1$ (see Corollary 3.2 below), defined in (3.3), and provide a uniform bound of the segment process $(Y_t^\varepsilon)_{t\in[0,T]}$ (see Lemma 4.3 below). Next, (**A4**) ensures that the displacement of the segment process $(X_t^\varepsilon)_{t\in[0,T]}$ is continuous in the mean L^p-norm sense (see Lemma 4.1 below).

3 Ergodicity of the Frozen Equation with Memory

Consider an SDE with memory associated with the fast motion while with the frozen slow component in the form

(3.1) $\mathrm{d}Y(t) = b_2(\zeta, Y(t), Y(t-\tau))\mathrm{d}t + \sigma_2(\zeta, Y(t), Y(t-\tau))\mathrm{d}W_2(t), \quad t > 0, \quad Y_0 = \eta \in \mathscr{C}.$

Under (**A2**), (3.1) has a unique strong solution $(Y(t))_{t\ge -\tau}$ (see, e.g., [24, Theorem 2.2, P.150]). To highlight the initial value $\eta \in \mathscr{C}$ and the frozen segment $\zeta \in \mathscr{C}$, we write the corresponding solution process $(Y^\zeta(t, \eta))_{t\ge -\tau}$ and the segment process $(Y_t^\zeta(\eta))_{t\ge 0}$ instead of $(Y(t))_{t\ge -\tau}$ and $(Y_t)_{t\ge 0}$, respectively.

Our main result in this section is stated as below. It is concerned with ergodicity of the frozen SDE with memory.

Theorem 3.1. Under (**A2**) and (**A3**), $Y_t^\zeta(\eta)$ has a unique invariant measure μ^ζ, and there exists $\lambda > 0$ such that

$$|\mathbb{E}b_1(\zeta, Y_t^\zeta(\eta)) - \overline{b}_1(\zeta)| \lesssim \mathrm{e}^{-\lambda t}(1 + \|\eta\|_\infty + \|\zeta\|_\infty), \quad t \geq 0, \ \ \eta \in \mathscr{C}, \tag{3.2}$$

where

$$\overline{b}_1(\zeta) := \int_{\mathscr{C}} b_1(\zeta, \varphi)\mu^\zeta(\mathrm{d}\varphi), \quad \zeta \in \mathscr{C}. \tag{3.3}$$

Proof. The main idea of the proof concerning existence of an invariant measure goes back to [1, Lemma 2.4], which, nevertheless, involves functional SDEs with additive noises.

Let $\mathscr{P}(\mathscr{C})$ be the set of all probability measures on $\mathscr{C}$. W_2 denotes the L^2-Wasserstein distance on $\mathscr{P}(\mathscr{C})$ induced by the bounded distance $\rho(\xi, \eta) := 1 \wedge \|\xi - \eta\|_\infty$, i.e.,

$$W_2(\mu_1, \mu_2) = \inf_{\pi \in \mathscr{C}(\mu_1, \mu_2)} \big(\pi(\rho^2)\big)^{\frac{1}{2}}, \quad \mu_1, \mu_2 \in \mathscr{P}(\mathscr{C}),$$

where $\mathscr{C}(\mu_1, \mu_2)$ is the set of all coupling probability measures with marginals μ_1 and μ_2. It is well known that $\mathscr{P}(\mathscr{C})$ is a complete metric space w.r.t. the distance W_2 ([7, Lemma 5.3, P.174] and [7, Theorem 5.4, P.175]), and the convergence in W_2 is equivalent to the weak convergence ([7, Theorem 5.6, P.179]). Let $P_t^{\zeta,\eta}$ be the law of the segment process $Y_t^\zeta(\eta)$. According to the Krylov-Bogoliubov existence theorem ([9, Theorem 3.1.1, P.21]), if $P_t^{\zeta,\eta}$ converges weakly to a probability measure μ_η^ζ, then μ_η^ζ is an invariant measure. So, in light of the previous discussion, it suffices to prove the assertions below:

(i) $\{P_t^{\zeta,\eta}\}_{t\geq 0}$ is a Cauchy sequence w.r.t. the distance W_2. If so, by the completeness of $\mathscr{P}(\mathscr{C})$ w.r.t. the distance W_2, there is $\mu_\eta^\zeta \in \mathscr{P}(\mathscr{C})$ such that $\lim_{t\to\infty} W_2(P_t^{\zeta,\eta}, \mu_\eta^\zeta) = 0$;

(ii) $W_2(\mu_\eta^\zeta, \mu_{\eta'}^\zeta) = 0$ for any $\eta, \eta' \in \mathscr{C}$ and frozen $\zeta \in \mathscr{C}$, that is, μ_η^ζ is independent of η.

In the sequel, we shall claim that (i) and (ii) hold, respectively. For any $t_2 > t_1 > \tau$ and the frozen segment $\zeta \in \mathscr{C}$, consider the following SDE with memory

$$\mathrm{d}\overline{Y}(t) = b_2(\zeta, \overline{Y}(t), \overline{Y}(t-\tau))\mathrm{d}t + \sigma_2(\zeta, \overline{Y}(t), \overline{Y}(t-\tau))\mathrm{d}W_2(t), \ t \in [t_2 - t_1, t_2] \tag{3.4}$$

with the initial value $\overline{Y}_{t_2-t_1} = \eta$. The solution process and the segment process associated with (3.4) are denoted by $(\overline{Y}^\zeta(t,\eta))$ and $(Y_t^\zeta(\eta))$, respectively. Observe that the laws of $Y_{t_2}^\zeta(\eta)$ and $\overline{Y}_{t_2}^\zeta(\eta)$ are $P_{t_2}^{\zeta,\eta}$ and $P_{t_1}^{\zeta,\eta}$, respectively.

By (**A2**), there exists an $\alpha > 0$ such that

$$\|\sigma_2(\chi, x, y) - \sigma_2(\chi, x', y')\| \le \alpha(|x - x'| + |y - y'|), \tag{3.5}$$

and

$$|b_2(\chi, 0, 0)| + \|\sigma_2(\chi, 0, 0)\| \le \alpha(1 + \|\chi\|_\infty) \tag{3.6}$$

for any $\chi \in \mathscr{C}$ and $x, x', y, y' \in \mathbb{R}^n$. Accordingly, (3.5) and (3.6), together with (**A3**), yield that there exist $\lambda_1' > \lambda_2' > 0$, independent of χ, such that

$$2\langle x, b_2(\chi, x, y)\rangle + \|\sigma_2(\chi, x, y)\|^2 \le -\lambda_1'|x|^2 + \lambda_2'|y|^2 + c(1 + \|\chi\|_\infty^2) \tag{3.7}$$

for any $\chi \in \mathscr{C}$ and $x, y \in \mathbb{R}^n$. For a sufficiently small $\lambda' > 0$ obeying $\lambda_1' - \lambda' - \lambda_2' \mathrm{e}^{\lambda'\tau} = 0$ due to $\lambda_1' > \lambda_2' > 0$, applying Itô's formula, we infer from (3.7) that

$$\begin{aligned}
\mathrm{e}^{\lambda' t}\mathbb{E}|Y^\zeta(t, \eta)|^2 &\le |\eta(0)|^2 + \int_0^t \mathrm{e}^{\lambda' s}\mathbb{E}\{c(1 + \|\zeta\|_\infty^2) + \lambda'|Y^\zeta(s, \eta)|^2 \\
&\qquad - \lambda_1'|Y^\zeta(s, \eta)|^2 + \lambda_2'|Y^\zeta(s - \tau, \eta)|^2\}\mathrm{d}s \\
&\lesssim \|\eta\|_\infty^2 + \mathrm{e}^{\lambda' t}(1 + \|\zeta\|_\infty^2), \quad t > 0.
\end{aligned}$$

Consequently, we arrive at

$$\mathbb{E}|Y^\zeta(t, \eta)|^2 \lesssim \mathrm{e}^{-\lambda' t}\|\eta\|_\infty^2 + 1 + \|\zeta\|_\infty^2, \quad t > 0. \tag{3.8}$$

Also, by the Itô formula, in addition to the Burkhold-Davis-Gundy (B-D-G for abbreviation) inequality, we derive from (**A3**), and (3.5)-(3.8) that, for any $t \ge \tau$,

$$\begin{aligned}
&\mathbb{E}\|Y_t^\zeta(\eta)\|_\infty^2 \\
&\lesssim 1 + \|\zeta\|_\infty^2 + \mathbb{E}|Y^\zeta(t - \tau, \eta)|^2 + \int_{t-2\tau}^t \mathbb{E}|Y^\zeta(s, \eta)|^2\mathrm{d}s \\
&\quad + 2\mathbb{E}\Big(\sup_{t-\tau \le s \le t}\Big|\int_{t-\tau}^s \langle Y^\zeta(s, \eta), \sigma_2(\zeta, Y^\zeta(s, \eta), Y^\zeta(s - \tau, \eta))\mathrm{d}W_2(s)\rangle\Big|\Big) \\
&\le \frac{1}{2}\mathbb{E}\|Y_t^\zeta(\eta)\|_\infty^2 + c\Big\{1 + \|\zeta\|_\infty^2 + \mathbb{E}|Y^\zeta(t - \tau, \eta)|^2 + \int_{t-2\tau}^t \mathbb{E}|Y^\zeta(s, \eta)|^2\mathrm{d}s\Big\}.
\end{aligned} \tag{3.9}$$

On the other hand, following the argument leading to (3.9), one has

$$\mathbb{E}\|Y_t^\zeta(\eta)\|_\infty^2 \le \frac{1}{2}\mathbb{E}\|Y_t^\zeta(\eta)\|_\infty^2 + c\Big\{1 + \|\zeta\|_\infty^2 + \|\eta\|_\infty^2 + \int_0^t \mathbb{E}|Y^\zeta(s, \eta)|^2\mathrm{d}s\Big\}, \ t \in [0, \tau]. \tag{3.10}$$

Thus, combining (3.8) with (3.9) and (3.10) leads to

$$\mathbb{E}\|Y_t^\zeta(\eta)\|_\infty^2 \le c(\mathrm{e}^{-\lambda' t}\|\eta\|_\infty^2 + 1 + \|\zeta\|_\infty^2). \tag{3.11}$$

In what follows, we assume $t \in [t_2 - t_1, t_2]$, and set $\Gamma^\zeta(t,\eta) := Y^\zeta(t,\eta) - \overline{Y}^\zeta(t,\eta)$ for the sake of notational simplicity. Again, for a sufficiently small $\lambda > 0$ such that $\lambda_1 - \lambda - \lambda_2 \mathrm{e}^{\lambda\tau} = 0$ owing to $\lambda_1 > \lambda_2$, by the Itô formula, it follows from (**A2**) that

$$\begin{aligned}
\mathrm{e}^{\lambda t}\mathbb{E}|\Gamma^\zeta(t,\eta)|^2 &\le \mathrm{e}^{\lambda(t_2-t_1)}\mathbb{E}|\Gamma^\zeta(t_2-t_1,\eta)|^2 \\
&\qquad + \int_{t_2-t_1}^t \mathrm{e}^{\lambda s}\mathbb{E}\{(\lambda-\lambda_1)|\Gamma^\zeta(s,\eta)|^2 + \lambda_2|\Gamma^\zeta(s-\tau,\eta)|^2\}\mathrm{d}s \\
&\le \mathrm{e}^{\lambda(t_2-t_1)}\mathbb{E}|\Gamma^\zeta(t_2-t_1,\eta)|^2 + \mathrm{e}^{\lambda\tau}\int_{t_2-t_1-\tau}^{t_2-t_1} \mathrm{e}^{\lambda s}\mathbb{E}|\Gamma^\zeta(s,\eta)|^2\mathrm{d}s \\
&\lesssim \mathrm{e}^{\lambda(t_2-t_1)}\|\eta\|_\infty^2 + \mathrm{e}^{\lambda(t_2-t_1)}\mathbb{E}\|Y_{t_2-t_1}^\zeta(\eta)\|_\infty^2.
\end{aligned}$$

This, together with (3.11), yields that

$$\mathbb{E}|\Gamma^\zeta(t,\eta)|^2 \lesssim \mathrm{e}^{-\lambda(t+t_1-t_2)}(1 + \|\eta\|_\infty^2 + \|\zeta\|_\infty^2). \tag{3.12}$$

Imitating a similar procedure to derive (3.9), in particular, we obtain from (**A2**), (3.5), and (3.12) that

$$\mathbb{E}\|\Gamma_{t_2}^\zeta(\eta)\|_\infty^2 \lesssim \mathrm{e}^{-\lambda t_1}(1 + \|\eta\|_\infty^2 + \|\zeta\|_\infty^2). \tag{3.13}$$

This further implies that

$$W_2(P_{t_1}^{\zeta,\eta}, P_{t_2}^{\zeta,\eta}) \le \mathbb{E}\{1 \wedge \|Y_{t_2}^\zeta(\eta) - \overline{Y}_{t_2}^\zeta(\eta)\|_\infty^2\} \lesssim \mathrm{e}^{-\lambda t_1}(1 + \|\eta\|_\infty^2 + \|\zeta\|_\infty^2),$$

which goes to zero as t_1 (hence t_2) tends to ∞. Thus claim (i) holds.

By carrying out a similar argument to obtain (3.13), one finds that

$$\mathbb{E}\|Y_t^\zeta(\eta) - Y_t^\zeta(\eta')\|_\infty^2 \lesssim \mathrm{e}^{-\lambda t}\|\eta - \eta'\|_\infty^2. \tag{3.14}$$

For fixed $\zeta \in \mathscr{C}$ and arbitrary $\eta, \eta \in \mathscr{C}$, observe that

$$W_2(\mu_\eta^\zeta, \mu_{\eta'}^\zeta) \le W_2(P_t^{\zeta,\eta}, \mu_\eta^\zeta) + W_2(P_t^{\zeta,\eta'}, \mu_{\eta'}^\zeta) + W_2(P_t^{\zeta,\eta}, P_t^{\zeta,\eta'}). \tag{3.15}$$

Consequently, claim (ii) follows by taking (3.14) and (3.15) into consideration.

By virtue of (3.11) and the invariance of μ^ζ, it then follows that

$$\int_{\mathscr{C}} \|\psi\|_\infty^2 \pi^\zeta(\mathrm{d}\psi) \le c\Big\{1 + \|\zeta\|_\infty^2 + \mathrm{e}^{-\lambda t}\int_{\mathscr{C}} \|\psi\|_\infty^2 \pi^\zeta(\mathrm{d}\psi)\Big\}.$$

Thus, choosing $t > 0$ sufficiently large such that $\delta := c\mathrm{e}^{-\lambda t} < 1$, one finds that

$$\int_{\mathscr{C}} \|\psi\|_\infty^2 \pi^\zeta(\mathrm{d}\psi) \lesssim 1 + \|\zeta\|_\infty^2. \tag{3.16}$$

Next, with the aid of the invariance of π^ζ, (3.14), and (3.16), we deduce from (**A1**) that

$$\begin{aligned}|\mathbb{E}b_1(\zeta, Y_t^\zeta(\eta)) - \overline{b}_1(\zeta)| &\lesssim \int_{\mathscr{C}} \mathbb{E}\|Y_t^\zeta(\eta) - Y_t^\zeta(\psi)\|_\infty \pi^\zeta(\mathrm{d}\psi) \lesssim \mathrm{e}^{-\frac{\lambda t}{2}} \int_{\mathscr{C}} \|\eta - \psi\|_\infty \pi^\zeta(\mathrm{d}\psi) \\ &\lesssim \mathrm{e}^{-\frac{\lambda t}{2}}(1 + \|\eta\|_\infty + \|\zeta\|_\infty).\end{aligned}$$

As a result, (3.2) follows. □

Remark 3.1. It should be noted that there are other alternative approaches to obtain existence and uniqueness of invariant measures for functional SDEs. Regarding to existence of invariant measures, Es-Sarhir et al. [8], and Kinnally-Williams [17] by Arzelà–Ascoli's tightness characterization, Bao et al. [2] using a remote start method, Bao et al. [3] adopting Kurtz's Tightness Criterion, and Reiβ et al. [26] by considering the semi-martingale characteristics. As for uniqueness of invariant measures, we refer to Hairer et al. [15], and Kinnally-Williams [17] by utilizing an asymptotic coupling method.

The next corollary, which plays a crucial role in discussing strong limit theorem for the averaging principle, states that $\overline{b}_1$, defined by (3.3), enjoys a Lipschitz property.

Corollary 3.2. Under (**A1**)-(**A3**), $\overline{b}_1 : \mathscr{C} \mapsto \mathbb{R}^n$, defined as in (3.3), is Lipschitz.

Proof. For arbitrary $\phi, \zeta \in \mathscr{C}$, let

$$\nabla_\phi \overline{b}_1(\zeta) = \frac{\mathrm{d}}{\mathrm{d}\varepsilon} \overline{b}_1(\zeta + \varepsilon\phi)\Big|_{\varepsilon=0}$$

be the direction derivative of $\overline{b}_1$ at ζ along the direction ϕ. By Theorem 3.1, we have

$$\begin{aligned}\nabla_\phi \overline{b}_1(\zeta) &= \lim_{t\to\infty} \mathbb{E}\nabla_\phi b_1(\zeta, Y_t^\zeta(\eta)) \\ &= \lim_{t\to\infty} \mathbb{E}\Big\{(\nabla_\phi^{(1)} b_1)(\zeta, Y_t^\zeta(\eta)) + \Big(\nabla^{(2)}_{\nabla_\phi Y_t^\zeta(\eta)} b_1\Big)(\zeta, Y_t^\zeta(\eta))\Big\}, \quad \phi, \zeta, \eta \in \mathscr{C}.\end{aligned}$$

According to (**A1**), to verify that $\overline{b}_1 : \mathscr{C} \mapsto \mathbb{R}^n$ is Lipschitz, it remains to verify

$$\sup_{t\geq 0} \mathbb{E}\|\nabla_\phi Y_t^\zeta(\eta)\|_\infty^2 < \infty. \tag{3.17}$$

Observe that $\nabla_\phi Y^\zeta(t,\eta)$ satisfies the following linear SDE with memory

$$\begin{aligned}
\mathrm{d}(\nabla_\phi Y^\zeta(t,\eta)) = \Big\{&(\nabla_\phi^{(1)} b_2)(\zeta, Y^\zeta(t,\eta), Y^\zeta(t-\tau,\eta))\\
&+ \Big(\nabla^{(2)}_{\nabla_\phi Y^\zeta(t,\eta)} b_2\Big)(\zeta, Y^\zeta(t,\eta), Y^\zeta(t-\tau,\eta))\\
&+ \Big(\nabla^{(3)}_{\nabla_\phi Y^\zeta(t-\tau,\eta)} b_2\Big)(\zeta, Y^\zeta(t,\eta), Y^\zeta(t-\tau,\eta))\Big\}\mathrm{d}t\\
&+ \Big\{(\nabla_\phi^{(1)} \sigma_2)(\zeta, Y^\zeta(t,\eta), Y^\zeta(t-\tau,\eta))\\
&+ \Big(\nabla^{(2)}_{\nabla_\phi Y^\zeta(t,\eta)} \sigma_2\Big)(\zeta, Y^\zeta(t,\eta), Y^\zeta(t-\tau,\eta))\\
&+ \Big(\nabla^{(3)}_{\nabla_\phi Y^\zeta(t-\tau,\eta)} \sigma_2\Big)(\zeta, Y^\zeta(t,\eta), Y^\zeta(t-\tau,\eta))\Big\}\mathrm{d}W_2(t), \quad t>0
\end{aligned}$$

with the initial data $\nabla_\phi Y_0^\zeta(\eta) = 0$. In the sequel, let $\chi \in \mathscr{C}$ and $x, x', y, y' \in \mathbb{R}^n$. For any $\varepsilon > 0$, it is trivial to see from (**A3**) that

$$\begin{aligned}
&2\varepsilon\langle x, b_2(\chi, x'+\varepsilon x, y'+\varepsilon y) - b_2(\chi, x', y')\rangle + \|\sigma_2(\chi, x'+\varepsilon x, y'+\varepsilon y) - \sigma_2(\chi, x', y')\|^2\\
&\quad \le -\lambda_1\varepsilon^2|x|^2 + \lambda_2\varepsilon^2|y|^2.
\end{aligned}$$

Multiplying ε^{-2} on both sides, followed by sending $\varepsilon \downarrow 0$, gives that

$$\begin{aligned}
&2\langle x, (\nabla_x^{(2)} b_2)(\chi, x', y') + (\nabla_y^{(3)} b_2)(\chi, x', y')\rangle\\
&\qquad + \|(\nabla_x^{(2)}\sigma_2)(\chi, x', y') + (\nabla_y^{(3)}\sigma_2)(\chi, x', y')\|^2\\
&\le -\lambda_1|x|^2 + \lambda_2|y|^2.
\end{aligned} \tag{3.18}$$

On the other hand, by virtue of (3.5), for any $\varepsilon > 0$, one has

$$\|\sigma_2(\chi, x'+\varepsilon x, y'+\varepsilon y) - \sigma_2(\chi, x', y')\|^2 \le \alpha\varepsilon^2(|x|^2 + |y|^2),$$

which further yields that

$$\|(\nabla_x^{(2)}\sigma_2)(\chi, x', y') + (\nabla_y^{(3)}\sigma_2)(\chi, x', y')\|^2 \le \alpha(|x|^2 + |y|^2). \tag{3.19}$$

Thus, with (3.18) and (3.19) in hand, (3.17) holds by repeating the argument which (3.11) is obtained. □

4 Preliminary Results

In this paper, we study the strong deviation between the slow component $X^\varepsilon(t)$ and the averaged component $\overline{X}(t)$, which satisfies the following functional SDE

$$\mathrm{d}\overline{X}(t) = \overline{b}_1(\overline{X}_t)\mathrm{d}t + \sigma_1(\overline{X}_t)\mathrm{d}W_1(t), \quad \overline{X}_0 = \xi \in \mathscr{C}, \tag{4.1}$$

where $\overline{b}_1 : \mathscr{C} \mapsto \mathbb{R}^n$ is defined as in (3.3). To achieve this goal, we need to construct some auxiliary two-time-scale stochastic systems with memory and provide a number of preliminary lemmas.

Throughout this paper, we fix $T > 0$ and set $\delta := \frac{\tau}{N} \in (0, 1)$ for a positive integer N sufficiently large. For any $t \in [0, T]$, consider the following auxiliary two-time-scale systems of functional SDEs

$$\mathrm{d}\widetilde{X}^\varepsilon(t) = b_1(X^\varepsilon_{t_\delta}, \widetilde{Y}^\varepsilon_t)\mathrm{d}t + \sigma_1(X^\varepsilon_{t_\delta})\mathrm{d}W_1(t), \quad X^\varepsilon_0 = \xi \in \mathscr{C}, \tag{4.2}$$

and

$$\begin{cases} \mathrm{d}\widetilde{Y}^\varepsilon(t) = \frac{1}{\varepsilon} b_2(X^\varepsilon_{t_\delta}, \widetilde{Y}^\varepsilon(t), \widetilde{Y}^\varepsilon(t-\tau))\mathrm{d}t + \frac{1}{\sqrt{\varepsilon}}\sigma_2(X^\varepsilon_{t_\delta}, \widetilde{Y}^\varepsilon(t), \widetilde{Y}^\varepsilon(t-\tau))\mathrm{d}W_2(t), \\ \widetilde{Y}^\varepsilon(t_\delta) = Y^\varepsilon(t_\delta) \end{cases} \tag{4.3}$$

with the initial value $\widetilde{Y}^\varepsilon_0 = Y^\varepsilon_0 = \eta \in \mathscr{C}$, where $t_\delta := \lfloor t/\delta \rfloor \delta$, the nearest breakpoint preceding t, with $\lfloor t/\delta \rfloor$ being the integer part of t/δ.

To proceed, we present several preliminary lemmas. The first lemma concerns the continuity in the mean L^p-norm sense for the displacement of the segment process $(X^\varepsilon_t)_{t\in[0,T]}$.

Lemma 4.1. Under (**A1**) and (**A4**),

$$\sup_{t\in[0,T]} \mathbb{E}\|X^\varepsilon_t - X^\varepsilon_{t_\delta}\|^p_\infty \lesssim_T \delta^{\frac{p-2}{2}}, \quad p > 2.$$

Proof. In accordance with [24, Theorem 4.1, P.160], we have

$$\mathbb{E}\Big(\sup_{0\le t\le T} \|X^\varepsilon_t\|^p_\infty\Big) \lesssim_T 1 + \|\xi\|^p_\infty. \tag{4.4}$$

Observe that

$$\begin{aligned} \mathbb{E}\|X^\varepsilon_t - X^\varepsilon_{t_\delta}\|^p_\infty &\le \sum_{m=0}^{N-1} \mathbb{E}\Big(\sup_{-(m+1)\delta\le\theta\le -m\delta} |X^\varepsilon(t+\theta) - X^\varepsilon(t_\delta+\theta)|^p\Big) \\ &=: \sum_{m=0}^{N-1} J_p(t, m, \delta), \end{aligned}$$

where $N = \tau/\delta$ by the definition of δ. To complete the proof of Lemma 4.1, it is sufficient to show

$$J_p(t, m, \delta) \lesssim_T \delta^{\frac{p}{2}}. \tag{4.5}$$

For any $t \in [0,T]$, take $k \geq 0$ such that $t \in [k\delta,(k+1)\delta)$. Thus, for any $\theta \in [-(m+1)\delta, -m\delta]$, one has

$$t+\theta \in [(k-m-1)\delta,(k+1-m)\delta] \ \text{ and } \ t_\delta+\theta \in [(k-m-1)\delta,(k-m)\delta].$$

In what follows, we consider three cases.

Case 1: $m \leq k-1$. Invoking Hölder's inequality and B-D-G's inequality, we obtain from (**A1**) and (4.4) that

$$\begin{aligned}
&J_p(t,m,\delta)\\
&\lesssim \delta^{p-1}\int_{(k-m-1)\delta}^{t-m\delta}\mathbb{E}|b_1(X_s^\varepsilon,Y_s^\varepsilon)|^p\mathrm{d}s+\mathbb{E}\Big(\sup_{-(m+1)\delta\leq\theta\leq-m\delta}\Big|\int_{k\delta+\theta}^{t+\theta}\sigma_1(X_s^\varepsilon)\mathrm{d}W_1(s)\Big|^p\Big)\\
&\lesssim \delta^{p-1}\int_{(k-m-1)\delta}^{t-m\delta}\mathbb{E}|b_1(X_s^\varepsilon,Y_s^\varepsilon)|^p\mathrm{d}s+\mathbb{E}\Big(\Big|\int_{(k-m-1)\delta}^{t-(m+1)\delta}\sigma_1(X_s^\varepsilon)\mathrm{d}W_1(s)\Big|^p\Big)\\
&\quad+\mathbb{E}\Big(\sup_{-(m+1)\delta\leq\theta\leq-m\delta}\Big|\int_{t-(m+1)\delta}^{t+\theta}\sigma_1(X_s^\varepsilon)\mathrm{d}W_1(s)\Big|^p\Big)\\
&\quad+\mathbb{E}\Big(\sup_{-(m+1)\delta\leq\theta\leq-m\delta}\Big|\int_{(k-m-1)\delta}^{k\delta+\theta}\sigma_1(X_s^\varepsilon)\mathrm{d}W_1(s)\Big|^p\Big)\\
&\lesssim \delta^{p-1}\int_{(k-m-1)\delta}^{t-m\delta}\mathbb{E}|b_1(X_s^\varepsilon,Y_s^\varepsilon)|^p\mathrm{d}s+\delta^{\frac{p-2}{2}}\mathbb{E}\Big(\int_{(k-m-1)\delta}^{t-(m+1)\delta}\|\sigma_1(X_s^\varepsilon)\|^p\mathrm{d}s\Big)\\
&\quad+\mathbb{E}\Big(\int_{t-(m+1)\delta}^{t-m\delta}\|\sigma_1(X_s^\varepsilon)\|^2\mathrm{d}s\Big)^{p/2}+\mathbb{E}\Big(\int_{(k-m-1)\delta}^{(k-m)\delta}\|\sigma_1(X_s^\varepsilon)\|^2\mathrm{d}s\Big)^{p/2}\\
&\lesssim_T \delta^{\frac{p}{2}}.
\end{aligned}\tag{4.6}$$

Case 2: $m \geq k+1$. In view of (**A5**), it follows that

$$|X^\varepsilon(t+\theta)-X^\varepsilon(t_\delta+\theta)|^p=|\xi(t+\theta)-\xi(t_\delta+\theta)|^p\lesssim\delta^p.$$

Case 3: $m=k$. Also, by Hölder's inequality and B-D-G's inequality, we deduce from (**A1**)

and (4.4) that

$$
\begin{aligned}
J_p(t,m,\delta) &= \mathbb{E}\Big(\sup_{-(k+1)\delta\le\theta\le -k\delta}|X^\varepsilon(t+\theta)-X^\varepsilon(k\delta+\theta)|^p\Big)\\
&\lesssim \delta^p+\mathbb{E}\Big(\sup_{-(k+1)\delta\le\theta\le -k\delta}(|X^\varepsilon(t+\theta)-X^\varepsilon(0)|^p\mathbf{1}_{\{t+\theta>0\}})\Big)\\
&\lesssim \delta^p+\mathbb{E}\Big(\sup_{-t\le\theta\le -k\delta}\Big|\int_0^{t+\theta}b_1(X_s^\varepsilon,Y_s^\varepsilon)\mathrm{d}s\Big|^p\Big)\\
&\quad+\mathbb{E}\Big(\sup_{-t\le\theta\le -k\delta}\Big|\int_0^{t+\theta}\sigma_1(X_s^\varepsilon)\mathrm{d}W_1(s)\Big|^p\Big)\\
&\lesssim \delta^p+\delta^{p-1}\int_0^{t-k\delta}\mathbb{E}|b_1(X_s^\varepsilon,Y_s^\varepsilon)|^p\mathrm{d}s+\delta^{\frac{p-2}{2}}\int_0^{t-k\delta}\mathbb{E}\|\sigma_1(X_s^\varepsilon)\|^p\mathrm{d}s\\
&\lesssim_T \delta^{\frac{p}{2}},
\end{aligned}\tag{4.7}
$$

where $a^+:=\max\{a,0\}$ for $a\in\mathbb{R}$. Consequently, the desired assertion (4.5) is finished by taking the discussions above into account. □

The lemma below provides an error bound of the difference in the strong sense between the slow component $(X^\varepsilon(t))$ and its approximation $(\widetilde{X}^\varepsilon(t))$.

Lemma 4.2. Assume that (**A1**) and (**A2**) hold and suppose further $\varepsilon/\delta\in(0,1)$. Then, there exists $\beta>0$ such that

$$
\mathbb{E}\Big(\sup_{0\le s\le T}|X^\varepsilon(t)-\widetilde{X}^\varepsilon(t)|^p\Big)\lesssim_T \delta^{\frac{p-2}{2}}(1+\varepsilon^{-1}\mathrm{e}^{\frac{\beta\delta}{\varepsilon}}),\quad p>2.
$$

Proof. In view of Hölder's inequality and B-D-G's inequality, it follows from (**A1**) and Lemma 4.1 that

$$
\begin{aligned}
\mathbb{E}\Big(\sup_{0\le s\le t}|X^\varepsilon(s)-\widetilde{X}^\varepsilon(s)|^p\Big)&\lesssim_T\int_0^t\mathbb{E}\{\|X_s^\varepsilon-X_{s_\delta}^\varepsilon\|_\infty^p+\|Y_s^\varepsilon-\widetilde{Y}_s^\varepsilon\|_\infty^p\}\mathrm{d}s\\
&\lesssim_T \delta^{\frac{p-2}{2}}+\int_0^t\mathbb{E}\|Y_s^\varepsilon-\widetilde{Y}_s^\varepsilon\|_\infty^p\mathrm{d}s,\quad t\in(0,T].
\end{aligned}
$$

Therefore, to finish the argument of Lemma 4.2, it suffices to show that there exists $\beta>0$ such that

$$
\sup_{t\in[0,T]}\mathbb{E}\|Y_t^\varepsilon-\widetilde{Y}_t^\varepsilon\|_\infty^p\lesssim_T\varepsilon^{-1}\delta^{\frac{p-2}{2}}\mathrm{e}^{\frac{\beta\delta}{\varepsilon}}.\tag{4.8}
$$

In what follows, we verify claim (4.8) by an induction argument. For any $t\in[0,\tau)$, due to $Y_0^\varepsilon=\widetilde{Y}_0^\varepsilon=\eta$, it is readily to check that

$$
\mathbb{E}\|Y_t^\varepsilon-\widetilde{Y}_t^\varepsilon\|_\infty^p\le\sum_{j=0}^{\lfloor t/\delta\rfloor}\mathbb{E}\Big(\sup_{j\delta\le s\le((j+1)\delta)\wedge t}|Y^\varepsilon(s)-\widetilde{Y}^\varepsilon(s)|^p\Big)=:I(t,\delta).
$$

By means of Itô's formula and B-D-G's inequality, together with $\widetilde{Y}^\varepsilon(t_\delta)=Y^\varepsilon(t_\delta)$, we obtain from (**A2**) that

$$\begin{aligned}&\mathbb{E}\Big(\sup_{j\delta\le s\le((j+1)\delta)\wedge t}|Y^\varepsilon(s)-\widetilde{Y}^\varepsilon(s)|^p\Big)\\&\quad\le\frac{c}{\varepsilon}\int_{j\delta}^{((j+1)\delta)\wedge t}\{\mathbb{E}\|X_s^\varepsilon-X_{s_\delta}^\varepsilon\|_\infty^2+\mathbb{E}|Y^\varepsilon(s)-\widetilde{Y}^\varepsilon(s)|^p\}\mathrm{d}s\\&\qquad+\frac{1}{2}\mathbb{E}\Big(\sup_{j\delta\le s\le((j+1)\delta)\wedge t}|Y^\varepsilon(s)-\widetilde{Y}^\varepsilon(s)|^p\Big),\quad t\in[0,\tau].\end{aligned}$$

Consequently, we conclude that

$$\begin{aligned}I(t,\delta)&\lesssim\frac{1}{\varepsilon}\int_0^t\mathbb{E}\|X_s^\varepsilon-X_{s_\delta}^\varepsilon\|_\infty^2\mathrm{d}s+\frac{1}{\varepsilon}\int_0^\delta\sum_{j=0}^{\lfloor t/\delta\rfloor}\mathbb{E}\Big(\sup_{j\delta\le r\le((j\delta+s))\wedge t}|Y^\varepsilon(r)-\widetilde{Y}^\varepsilon(r)|^p\Big)\mathrm{d}s\\&\lesssim\frac{1}{\varepsilon}\int_0^t\mathbb{E}\|X_s^\varepsilon-X_{s_\delta}^\varepsilon\|_\infty^2\mathrm{d}s+\frac{1}{\varepsilon}\int_0^\delta I(t,s)\mathrm{d}s.\end{aligned}\tag{4.9}$$

This, combining Lemma 4.1 with Gronwall's inequality, gives that

$$\mathbb{E}\|Y_t^\varepsilon-\widetilde{Y}_t^\varepsilon\|_\infty^p\lesssim\varepsilon^{-1}\delta^{\frac{p-2}{2}}\mathrm{e}^{\frac{c\delta}{\varepsilon}},\qquad t\in[0,\tau)\tag{4.10}$$

for some $c>0$. Next, for any $t\in[\tau,2\tau)$, thanks to (4.10), it is immediate to note that

$$\begin{aligned}\mathbb{E}\|Y_t^\varepsilon-\widetilde{Y}_t^\varepsilon\|_\infty^p&\le\mathbb{E}\Big(\|Y_\tau^\varepsilon-\widetilde{Y}_\tau^\varepsilon\|_\infty^p\Big)+\mathbb{E}\Big(\sup_{\tau\le s\le t}|Y^\varepsilon(s)-\widetilde{Y}^\varepsilon(s)|^p\Big)\\&\le c\Big\{\varepsilon^{-1}\delta^{\frac{p-2}{2}}\mathrm{e}^{\frac{c\delta}{\varepsilon}}+\sum_{j=0}^{\lfloor t-\tau\rfloor}\mathbb{E}\Big(\sup_{(N+j)\delta\le s\le((N+j+1)\delta)\wedge t}|Y^\varepsilon(s)-\widetilde{Y}^\varepsilon(s)|^p\Big)\Big\}\\&=:c\{\varepsilon^{-1}\delta^{\frac{p-2}{2}}\mathrm{e}^{\frac{c\delta}{\varepsilon}}+M(t,\tau,\delta)\}.\end{aligned}$$

Carrying out a similar argument to derive (4.9), we deduce from (4.10) that

$$\begin{aligned}M(t,\tau,\delta)&\lesssim\frac{1}{\varepsilon}\int_\tau^t\mathbb{E}\|X_s^\varepsilon-X_{s_\delta}^\varepsilon\|_\infty^2\mathrm{d}s\\&\quad+\frac{1}{\varepsilon}\int_0^\delta\sum_{j=0}^{\lfloor t-\tau\rfloor}\mathbb{E}\Big(\sup_{(N+j)\delta\le r\le((N+j)\delta+s)\wedge t}|Y^\varepsilon(r)-\widetilde{Y}^\varepsilon(r)|^p\Big)\mathrm{d}s\\&\quad+\frac{1}{\varepsilon}\int_0^\delta\sum_{j=0}^{\lfloor t-\tau\rfloor}\mathbb{E}\Big(\sup_{j\delta\le s\le((j+1)\delta)\wedge(t-\tau)}|Y^\varepsilon(s)-\widetilde{Y}^\varepsilon(s)|^p\Big)\mathrm{d}s\\&\lesssim\frac{\delta^{\frac{p-2}{2}}}{\varepsilon}+\frac{\delta}{\varepsilon}\cdot\frac{\delta^{\frac{p-2}{2}}}{\varepsilon}\mathrm{e}^{\frac{c\delta}{\varepsilon}}+\frac{1}{\varepsilon}\int_0^\delta M(t,\tau,s)\mathrm{d}s.\end{aligned}$$

Thus, the Gronwall inequality reads

$$M(t,\tau,\delta) \lesssim \Big\{\frac{\delta^{\frac{p-2}{2}}}{\varepsilon}+\frac{\delta}{\varepsilon}\cdot\frac{\delta^{\frac{p-2}{2}}}{\varepsilon}\mathrm{e}^{\frac{c\delta}{\varepsilon}}\Big\}\mathrm{e}^{\frac{c\delta}{\varepsilon}} \lesssim \frac{\delta}{\varepsilon}\cdot\frac{\delta^{\frac{p-2}{2}}}{\varepsilon}\mathrm{e}^{\frac{c\delta}{\varepsilon}} \lesssim \frac{\delta^{\frac{p-2}{2}}}{\varepsilon}\mathrm{e}^{\frac{c\delta}{\varepsilon}},$$

where we have used $\varepsilon/\delta \in (0,1)$ in the second step. Finally, (4.8) follows by repeating the previous procedure. □

The following consequence explores a uniform estimate w.r.t. the parameter ε for the segment process associated with the auxiliary fast motion.

Lemma 4.3. Assume that (**A1**) and (**A3**) hold. Then, there exists $C_T > 0$, independent of ε, such that

$$\sup_{t\in[0,T]} \mathbb{E}\|\widetilde{Y}_t^\varepsilon\|_\infty^2 \le C_T. \tag{4.11}$$

Proof. From (2.2), it follows that

$$\begin{aligned} Y^\varepsilon(t) = \eta(0) &+ \int_0^{t/\varepsilon} b_2(X_{\varepsilon s}^\varepsilon, Y^\varepsilon(\varepsilon s), Y^\varepsilon(\varepsilon s-\tau))\mathrm{d}t \\ &+ \int_0^{t/\varepsilon} \sigma_2(X_{\varepsilon s}^\varepsilon, Y^\varepsilon(\varepsilon s), Y^\varepsilon(\varepsilon s-\tau))\mathrm{d}\overline{W}_2(s), \quad t>0, \end{aligned} \tag{4.12}$$

where we used the fact that $\overline{W}(t) := \frac{1}{\sqrt{\varepsilon}}W_2(\varepsilon t)$ is a Brownian motion. For fixed $\varepsilon > 0$ and $t \ge 0$, let $\overline{Y}^\varepsilon(t+\theta) = Y^\varepsilon(\varepsilon t+\theta), \theta \in [-\tau, 0]$. So, one has $\overline{Y}_t^\varepsilon = Y_{\varepsilon t}^\varepsilon$. Observe that (4.12) can be rewritten as

$$\overline{Y}^\varepsilon(t/\varepsilon) = \eta(0) + \int_0^{t/\varepsilon} b_2(X_{\varepsilon s}^\varepsilon, \overline{Y}^\varepsilon(s), \overline{Y}^\varepsilon(s-\tau))\mathrm{d}s + \int_0^{t/\varepsilon} \sigma_2(X_{\varepsilon s}^\varepsilon, \overline{Y}^\varepsilon(s), \overline{Y}^\varepsilon(s-\tau))\mathrm{d}\overline{W}_2(s).$$

Then, following the argument to obtain (3.11), for any $s > 0$ we can deduce that

$$\mathbb{E}\|\overline{Y}_s^\varepsilon\|_\infty^2 \lesssim 1 + \|\eta\|_\infty^2\mathrm{e}^{-\lambda s} + \mathbb{E}\Big(\sup_{0\le r\le \varepsilon s}\|X_r^\varepsilon\|_\infty^2\Big).$$

This, together with $\overline{Y}_t^\varepsilon = Y_{\varepsilon t}^\varepsilon$, gives that

$$\mathbb{E}\|Y_{\varepsilon s}^\varepsilon\|_\infty^2 \lesssim 1 + \|\eta\|_\infty^2\mathrm{e}^{-\lambda s} + \mathbb{E}\Big(\sup_{0\le r\le \varepsilon s}\|X_r^\varepsilon\|_\infty^2\Big).$$

In particular, taking $s = t/\varepsilon$ we arrive at

$$\mathbb{E}\|Y_t^\varepsilon\|_\infty^2 \lesssim 1 + \|\eta\|_\infty^2 + \mathbb{E}\Big(\sup_{0\le r\le t}\|X_r^\varepsilon\|_\infty^2\Big).$$

This, together with (4.4), yields that

$$\sup_{t\in[0,T]} \mathbb{E}\|Y_t^\varepsilon\|_\infty^2 \le C_T$$

for some $C_T > 0$. Observe from (4.8) and Höder's inequality that

$$\begin{aligned}\mathbb{E}\|\widetilde{Y}_t^\varepsilon\|_\infty^2 &\le 2\mathbb{E}\|Y_t^\varepsilon - \widetilde{Y}_t^\varepsilon\|_\infty^2 + 2\mathbb{E}\|Y_t^\varepsilon\|_\infty^2 \\ &\lesssim_T 1 + \Big(\varepsilon^{-1}\delta^{\frac{p-2}{2}}\mathrm{e}^{\frac{\beta\delta}{\varepsilon}}\Big)^{2/p}, \quad p > 4.\end{aligned}$$

Next, taking $\delta = \varepsilon(-\ln\varepsilon)^{\frac{1}{2}}$ in the estimate above and letting $y = (-\ln\varepsilon)^{\frac{1}{2}}$, we have

$$\mathbb{E}\|\widetilde{Y}_t^\varepsilon\|_\infty^2 \lesssim_T 1 + \Big(\mathrm{e}^{y^2}(\mathrm{e}^{-y^2}y)^{\frac{p-2}{2}}\mathrm{e}^{\beta y}\Big)^{2/p}, \quad p > 4.$$

Then, the desired assertion follows since the leading term $\mathrm{e}^{y^2}(\mathrm{e}^{-y^2}y)^{\frac{p-2}{2}}\mathrm{e}^{\beta y} \to 0$ as $y \uparrow \infty$ whenever $p > 4$.

□

5 A Strong Limit Theorem for the Slow Component

With several preliminary lemmas at our hands, we are in position to present our main result.

Theorem 5.1. Under (**A1**)-(**A4**), one has

$$\lim_{\varepsilon\to 0}\mathbb{E}\Big(\sup_{0\le t\le T}|X^\varepsilon(t) - \overline{X}(t)|^p\Big) = 0, \quad p > 0.$$

Proof. For any $t \in [0, T]$ and $p > 0$, set

$$\varLambda(t) := \mathbb{E}\Big(\sup_{0\le s\le t}|X^\varepsilon(s) - \overline{X}(s)|^p\Big) \quad \text{and} \quad \varGamma(t) := \mathbb{E}\Big(\sup_{0\le s\le t}|\widetilde{X}^\varepsilon(s) - \overline{X}(s)|^p\Big).$$

By Hölder's inequality, it is sufficient to verify that

$$\lim_{\varepsilon\to 0}\varLambda(T) = 0, \quad p > 4. \tag{5.1}$$

In what follows, let $t \in [0, T]$ be arbitrary and assume $p > 4$. For any $t \in [0, T]$, it follows from Lemma 4.2 that

$$\varLambda(t) \lesssim \mathbb{E}\Big(\sup_{0\le s\le t}|X^\varepsilon(s) - \widetilde{X}^\varepsilon(s)|^p\Big) + \varGamma(t) \lesssim \delta^{\frac{p-2}{2}}\Big(1 + \frac{1}{\varepsilon}\mathrm{e}^{\frac{\beta\delta}{\varepsilon}}\Big) + \varGamma(t). \tag{5.2}$$

Next, if we can show that

$$\Gamma(t) \lesssim \delta^{\frac{p-2}{2}}\Big(1+\frac{1}{\varepsilon}\mathrm{e}^{\frac{\beta\delta}{\varepsilon}}\Big)+\Big(\frac{\varepsilon}{\delta}\Big)^{\nu}+\int_0^t \Lambda(s)\mathrm{d}s \tag{5.3}$$

for some $\nu \in (0,1)$, inserting (5.3) back into (5.2) and utilizing Gronwall's inequality, we deduce that

$$\Lambda(t) \lesssim \delta^{\frac{p-2}{2}}\Big(1+\frac{1}{\varepsilon}\mathrm{e}^{\frac{\beta\delta}{\varepsilon}}\Big)+\Big(\frac{\varepsilon}{\delta}\Big)^{\nu}.$$

Thus, the desired assertion (5.1) follows by choosing $\delta=\varepsilon(-\ln\varepsilon)^{\frac{1}{2}}$. Indeed, it is easy to see that $\varepsilon/\delta \in (0,1)$, which is prerequisite in Lemma 4.2, for $\varepsilon \in (0,1)$ small enough, and that $\delta \to 0$ as $\varepsilon \downarrow 0$. Furthermore, let $y=(-\ln\varepsilon)^{\frac{1}{2}}$ (hence $\varepsilon=\mathrm{e}^{-y^2}$), which goes into infinity as ε tends to zero. Then, we have

$$\Lambda(t) \lesssim (\mathrm{e}^{-y^2}y)^{\frac{p-2}{2}}\Big(1+\mathrm{e}^{y^2+\beta y}\Big)+y^{-\nu},$$

which goes to zero by taking $p>4$ and letting $y \uparrow \infty$.

Next, we intend to claim (5.3). Set

$$\Gamma_p(t,\delta,\varepsilon):=\mathbb{E}\Big(\sup_{0\le s\le t}\Big|\int_0^s\{b_1(X^\varepsilon_{r_\delta},\widetilde{Y}^\varepsilon_r)-\overline{b}_1(X^\varepsilon_{r_\delta})\}\mathrm{d}r\Big|^p\Big),\quad t\in[0,T].$$

Applying Hölder's inequality, B-D-G's inequality, Lipschitz property of $\overline{b}_1$ due to Corollary 3.2, and Lemma 4.1, we derive that

$$\begin{aligned}
\Gamma(t) &\lesssim \mathbb{E}\Big(\sup_{0\le s\le t}\Big|\int_0^t\{b_1(X^\varepsilon_{s_\delta},\widetilde{Y}^\varepsilon_s)-\overline{b}_1(\overline{X}_s)\}\mathrm{d}s\Big|^p\Big)+\int_0^t\mathbb{E}\|\sigma_1(X^\varepsilon_{s_\delta})-\sigma_1(\overline{X}_s)\|^p\mathrm{d}s\\
&\lesssim \Gamma_p(t,\delta,\varepsilon)+\int_0^t\mathbb{E}|\overline{b}_1(X^\varepsilon_{s_\delta})-\overline{b}_1(X^\varepsilon_s)|^p\mathrm{d}s+\int_0^t\mathbb{E}|\overline{b}_1(X^\varepsilon_s)-\overline{b}_1(\widetilde{X}^\varepsilon_s)|^p\mathrm{d}s\\
&\quad+\int_0^t\mathbb{E}|\overline{b}_1(\widetilde{X}^\varepsilon_s)-\overline{b}_1(\overline{X}_s)|^p\mathrm{d}s+\int_0^t\mathbb{E}\|\sigma_1(X^\varepsilon_{s_\delta})-\sigma_1(\overline{X}_s)\|^p\mathrm{d}s\\
&\lesssim \Gamma_p(t,\delta,\varepsilon)+\int_0^t\mathbb{E}\|X^\varepsilon_s-\widetilde{X}^\varepsilon_s\|_\infty\mathrm{d}s+\int_0^t\mathbb{E}\|X^\varepsilon_{s_\delta}-X^\varepsilon_s\|^p_\infty\mathrm{d}s+\int_0^t\Gamma(s)\mathrm{d}s+\int_0^t\Lambda(s)\mathrm{d}s\\
&\lesssim \delta^{\frac{p-2}{2}}+\frac{1}{\varepsilon}\delta^{\frac{p-2}{2}}\mathrm{e}^{\frac{c\delta}{\varepsilon}}+\Gamma_p(t,\delta,\varepsilon)+\int_0^t\Gamma(s)\mathrm{d}s+\int_0^t\Lambda(s)\mathrm{d}s,
\end{aligned}$$

which, together with Gronwall's inequality, leads to

$$\Gamma(t) \lesssim \delta^{\frac{p-2}{2}}\Big(1+\frac{1}{\varepsilon}\mathrm{e}^{\frac{\beta\delta}{\varepsilon}}\Big)+\Gamma_p(t,\delta,\varepsilon)+\int_0^t\Lambda(s)\mathrm{d}s, \tag{5.4}$$

where we have utilized the fact that $\varGamma_p(t,\delta,\varepsilon)$ is nondecreasing with respect to t. By comparing (5.3) with (5.4), we need only prove

$$\varGamma_p(t,\delta,\varepsilon) \lesssim \Big(\frac{\varepsilon}{\delta}\Big)^{\nu} \tag{5.5}$$

for some $\nu \in (0,1)$.

Let

$$\varUpsilon_p(k,\delta,\varepsilon) = \mathbb{E}\Big(\Big|\int_{k\delta}^{((k+1)\delta)\wedge t}\{b_1(X^\varepsilon_{k\delta},\widetilde{Y}^\varepsilon_s) - \overline{b}_1(X^\varepsilon_{k\delta})\}\mathrm{d}s\Big|^p\Big) \quad \text{for any } p>0.$$

In the sequel, we show that (5.5) holds. By Hölder's inequality, we obtain that

$$\begin{aligned}\varGamma_p(t,\delta,\varepsilon) &= \mathbb{E}\Big(\sup_{0\le s\le t}\Big|\sum_{k=0}^{\lfloor s/\delta\rfloor}\int_{k\delta}^{((k+1)\delta)\wedge t}\{b_1(X^\varepsilon_{k\delta},\widetilde{Y}^\varepsilon_r) - \overline{b}_1(X^\varepsilon_{k\delta})\}\mathrm{d}r\Big|^p\Big)\\ &\le \mathbb{E}\Big(\sup_{0\le s\le t}\Big((\lfloor s/\delta\rfloor+1)^{p-1}\sum_{k=0}^{\lfloor s/\delta\rfloor}\varUpsilon_p(k,\delta,\varepsilon)\Big)\Big)\\ &\le (\lfloor t/\delta\rfloor+1)^{p-1}\sum_{k=0}^{\lfloor t/\delta\rfloor}\varUpsilon_p(k,\delta,\varepsilon)\\ &\le (\lfloor t/\delta\rfloor+1)^{p}\max_{0\le k\le\lfloor t/\delta\rfloor}\varUpsilon_p(k,\delta,\varepsilon).\end{aligned}\tag{5.6}$$

For any $p'\in(1,2)$, by Hölder's inequality, (**A1**), and (4.4), observe that

$$\begin{aligned}\varUpsilon_p(k,\delta,\varepsilon) &\le \varUpsilon_2(k,\delta,\varepsilon)^{\frac{p'}{2}}\Big(\mathbb{E}\Big(\Big|\int_{k\delta}^{((k+1)\delta)\wedge t}\{b_1(X^\varepsilon_{k\delta},\widetilde{Y}^\varepsilon_s) - \overline{b}_1(X^\varepsilon_{k\delta})\}\mathrm{d}s\Big|^{\frac{2(p-p')}{2-p'}}\Big)\Big)^{\frac{2-p'}{2}}\\ &\le \varUpsilon_2(k,\delta,\varepsilon)^{\frac{p'}{2}}\Big(\delta^{\frac{2(p-p')}{2-p'}-1}\mathbb{E}\Big(\Big|\int_{k\delta}^{((k+1)\delta)\wedge t}|b_1(X^\varepsilon_{k\delta},\widetilde{Y}^\varepsilon_s) - \overline{b}_1(X^\varepsilon_{k\delta})|^{\frac{2(p-p')}{2-p'}}\mathrm{d}s\Big|\Big)\Big)^{\frac{2-p'}{2}}\\ &\lesssim \varUpsilon_2(k,\delta,\varepsilon)^{\frac{p'}{2}}\delta^{\frac{2(p-p')}{2-p'}\times\frac{2-p'}{2}}\\ &\lesssim \varUpsilon_2(k,\delta,\varepsilon)^{\frac{p'}{2}}\delta^{p-p'}, \qquad p>4.\end{aligned}$$

Substituting this into (5.6), we arrive at

$$\varGamma_p(t,\delta,\varepsilon) \lesssim \varUpsilon_2(k,\delta,\varepsilon)^{\frac{p'}{2}}\delta^{-p'}.$$

Thus, to complete the argument, it remains to show that

$$\varUpsilon_2(k,\delta,\varepsilon) \lesssim \varepsilon\delta.$$

Also, by virtue of Hölder's inequality, (**A1**), and (4.4), we derive that

$$
\begin{aligned}
&\Upsilon_2(k,\delta,\varepsilon)\\
&= 2\int_{k\delta}^{((k+1)\delta)\wedge t}\int_{s}^{((k+1)\delta)\wedge t}\mathbb{E}\langle b_1(X^\varepsilon_{k\delta},\widetilde{Y}^\varepsilon_s)-\overline{b}_1(X^\varepsilon_{k\delta}),b_1(X^\varepsilon_{k\delta},\widetilde{Y}^\varepsilon_r)-\overline{b}_1(X^\varepsilon_{k\delta})\rangle \mathrm{d}r\mathrm{d}s \qquad (5.7)\\
&\lesssim \int_{k\delta}^{(k+1)\delta}\int_{s}^{(k+1)\delta}(\mathbb{E}|\mathbb{E}((b_1(X^\varepsilon_{k\delta},\widetilde{Y}^\varepsilon_r)-\overline{b}_1(X^\varepsilon_{k\delta}))|\mathscr{F}_s)|^2)^{1/2}\mathrm{d}r\mathrm{d}s.
\end{aligned}
$$

For any $r\in[k\delta,(k+1)\delta)$, by the definition of $\widetilde{Y}^\varepsilon$, defined as in (4.3), it follows that

$$
\begin{aligned}
\widetilde{Y}^\varepsilon(r) &= \widetilde{Y}^\varepsilon(k\delta)+\frac{1}{\varepsilon}\int_{k\delta}^{r}b_2(X^\varepsilon_{k\delta},\widetilde{Y}^\varepsilon(u),\widetilde{Y}^\varepsilon(u-\tau))\mathrm{d}u\\
&\quad+\frac{1}{\sqrt{\varepsilon}}\int_{k\delta}^{r}\sigma_2(X^\varepsilon_{k\delta},\widetilde{Y}^\varepsilon(u),\widetilde{Y}^\varepsilon(u-\tau))\mathrm{d}W_2(u)\\
&= \widetilde{Y}^\varepsilon(k\delta)+\int_0^{\frac{r-k\delta}{\varepsilon}}b_2(X^\varepsilon_{k\delta},\widetilde{Y}^\varepsilon(k\delta+\varepsilon u),\widetilde{Y}^\varepsilon(k\delta+\varepsilon u-\tau))\mathrm{d}u \qquad (5.8)\\
&\quad+\int_0^{\frac{r-k\delta}{\varepsilon}}\sigma_2(X^\varepsilon_{k\delta},\widetilde{Y}^\varepsilon(k\delta+\varepsilon u-\tau))\mathrm{d}\widetilde{W}_2(u),
\end{aligned}
$$

where $\widetilde{W}_2(u):=(W_2(\varepsilon u+k\delta)-W(k\delta))/\sqrt{\varepsilon}$, which is also a Wiener process. For fixed $\varepsilon>0$ and $u\geq 0$, let

$$
\overline{Y}^{X^\varepsilon_{k\delta}}(u+\theta)=\widetilde{Y}^\varepsilon(k\delta+\varepsilon u+\theta),\quad \theta\in[-\tau,0].
$$

Then (5.8) can be rewritten as

$$
\begin{aligned}
\overline{Y}^{X^\varepsilon_{k\delta}}\Big(\frac{r-k\delta}{\varepsilon}\Big) &= \widetilde{Y}^\varepsilon(k\delta)+\int_0^{\frac{r-k\delta}{\varepsilon}}b_2\left(X^\varepsilon_{k\delta},\overline{Y}^{X^\varepsilon_{k\delta}}(u),\overline{Y}^{X^\varepsilon_{k\delta}}(u-\tau)\right)\mathrm{d}u\\
&\quad+\int_0^{\frac{r-k\delta}{\varepsilon}}\sigma_2\left(X^\varepsilon_{k\delta},\overline{Y}^{X^\varepsilon_{k\delta}}(u),\overline{Y}^{X^\varepsilon_{k\delta}}(u-\tau)\right)\mathrm{d}\widetilde{W}_2(u).
\end{aligned}
$$

Consequently, by the weak uniqueness of solution, we arrive at

$$
\mathscr{L}(\widetilde{Y}^\varepsilon_r)=\mathscr{L}\Big(Y^{X^\varepsilon_{k\delta}}_{(r-k\delta)/\varepsilon}(\widetilde{Y}^\varepsilon_{k\delta})\Big), \qquad (5.9)
$$

where $\mathscr{L}(\zeta)$ denotes the law of random variable ζ. Finally, we obtain from (3.2), (5.7), (5.9), and Lemma 4.3 that

$$
\begin{aligned}
\Upsilon_2(k,\delta,\varepsilon) &\lesssim (1+\mathbb{E}\|X^\varepsilon_{k\delta}\|^2_\infty+\mathbb{E}\|\widetilde{Y}^\varepsilon_{k\delta}\|^2_\infty)\int_{k\delta}^{(k+1)\delta}\int_{s}^{(k+1)\delta}\exp\Big(-\frac{c(r-k\delta)}{\varepsilon}\Big)\mathrm{d}r\mathrm{d}s\\
&\lesssim \varepsilon\delta.
\end{aligned}
$$

The proof is therefore complete.

□

Remark 5.1. In this paper, we only focus on the case, where the diffusion coefficient of the slow component is independent of the fast motion. For the case that the slow component fully depends on the fast one, there is an illustrative counterexample [22, p.1011] in which the weak convergence holds but there is no strong convergence.

Remark 5.2. In the present paper, we explore a strong limit theorem for the averaging principle for a class of two-time-scale SDEs with memory under certain dissipative conditions. Nevertheless, our main result can be generalized to some cases, where the fast motion does not satisfy a dissipative condition. Indeed, by a close inspection of the argument of Theorem 5.1, to cope with the non-dissipative case, one of the crucial procedures is to discuss the ergodic property of the frozen equation without dissipativity. However, for some special cases, this problem has been addressed in Bao et al. [3].

Remark 5.3. As we mentioned in the Introduction section, the study on two-time-scale stochastic systems with memory is still in its infancy. So, there is numerous work to be done in the future. Here, we list some of them. For the fast component, in this work we concentrate on the case of point delay. So far, it seems hard to extend our main result to the general case, e.g., the distributed delay, where the main difficulty is to provide an error bound of the difference in the strong sense between the fast component $(Y^\varepsilon(t))$ and its approximation $(\widetilde{Y}^\varepsilon(t))$. Moreover, it is also very challengeable to reveal the rate of strong convergence established in Theorem 5.1 since the phase space of the segment processes is infinite-dimensional. The questions above will be addressed in our forthcoming work.

References

[1] Bao, J., Wang, F.-Y., Yuan, C., Hypercontractivity for functional stochastic differential equations, *Stochastic Process. Appl.*, **125** (2015), 3636–3656.

[2] Bao, J., Yin, G., Yuan, C., Ergodicity for functional stochastic differential equations and applications, *Nonlinear Anal.*, **98** (2014), 66–82.

[3] Bao, J., Yin, G., Yuan, C., Stationary Distributions for Retarded Stochastic Differential Equations without Dissipativity, arXiv:1308.2018.

[4] Blömker, D., Hairer, M., Pavliotis, G. A., Multiscale analysis for stochastic partial differential equations with quadratic nonlineaities, *Nonlinearity*, **20** (2007), 1721–1744.

[5] Bréhier, C.-E., Strong and weak orders in averaging for SPDEs, *Stochastic Process. Appl.*, **122** (2012), 2553–2593.

[6] Cerrai, S., A Khasminskii type averaging principle for stochastic reaction-diffusion equations, *Ann. Appl. Probab.*, **19** (2009), 899–948.

[7] M.-F. Chen, *From Markov Chains to Non-Equilibrium Particle Systems,* Second Edition, World Scientific, Singapore, 2004.

[8] Es-Sarhir, A., Scheutzow, M., van Gaans, O., Invariant measures for stochastic functional differential equations with superlinear drift term, *Differential Integral Equations*, **23** (2010), 189–200.

[9] Da Prato, G., Zabczyk, J., *Ergodicity for infinite-dimensional systems*, In: London Mathematical Society. Lecture Note Series, vol. 229, Cambridge University Press, Cambridge, 1996.

[10] E, W., Liu, D., Vanden-Eijnden, E., Analysis of multiscale methods for stochastic differential equations, *Comm. Pure Appl. Math.*, **58** (2005), 1544–1585.

[11] Freidlin, M. I., Wentzell, A. D., *Random Perturbations of Dynamical Systems*, second ed., New York, Springer-Verlag, 1998.

[12] Fu, H., Wan, L., Liu, J., Strong convergence in averaging principle for stochastic hyperbolic-parabolic equations with two time-scales, *Stochastic Process. Appl.*, **125** (2015), 3255–3279.

[13] Givon, D., Strong convergence rate for two-time-scale jump-diffusion stochastic differential systems, *SIAM Multiscale Model. Simul.*, **6** (2007), 577–594.

[14] Givon, D., Kevrekidis, I. G., Kupferman, R., Strong convergence of projective integration schemes for singularly perturbed stochastic differential systems, *Commun. Math. Sci.*, **4** (2006), 707–729.

[15] Hairer, M., Mattingly, J. C., Scheutzow, M., Asymptotic coupling and a general form of Harris' theorem with applications to stochastic delay equations, *Probab. Theory Related Fields*, **149** (2011), 223–259.

[16] Kabanov, Y., Pergamenshchikov, S., *Two-Scale Stochastic Systems*, Springer, Berlin, 2003.

[17] Kinnally, M. S., Williams, R. J., On existence and uniqueness of stationary distributions for stochastic delay differential equations with positivity constraints, *Electron. J. Probab.*, **15** (2010), 409–451.

[18] Khasminskii, R. Z., On an averaging principle for Itô stochastic differential equations, *Kibernetica*, **4** (1968), 260-279.

[19] Kuksin, S. B., Piatnitski, A. L., Khasminski-Whitman averaging for randonly perturbed KdV equations, *J Math Pures Appl.*, **89** (2008), 400–428.

[20] Kushner, H. J. Large deviations for two-time-scale diffusions, with delays, *Appl. Math. Optim.*, **62** (2010), 295–322.

[21] Kushner, H. J., *Weak Convergence Methods and Singularly Perturbed Stochastic Control and Filtering Problems*, Birkhäuser, Boston, MA, 1990.

[22] Liu, D., Strong convergence of principle of averaging for multiscale stochastic dynamical systems, *Commun. Math. Sci.*, **8** (2010), 999–1020.

[23] Liu, Y., Yin, G., Asymptotic expansions of transition densities for hybrid jump-diffusions, *Acta Math. Appl. Sin. Engl. Ser.*, **20** (2004), 1–18.

[24] Mao, X., *Stochastic differential equations and applications*, 2nd Ed., Horwood Publishing Limited, Chichester, 2008.

[25] Mohammed, S-E. A., *Stochastic Functional Differential Equations*, Pitman, Boston, 1984.

[26] Reiβ, M., Riedle, M., van Gaans, O., Delay differential equations driven by Lvy processes: stationarity and Feller properties, *Stochastic Process. Appl.*, **116** (2006), 1409–1432.

[27] van Kampen, N. G., Elimination of fast variables, *Phys. Rep.,* **124** (1985), 69–160.

[28] Veretennikov, A. Y., On large deviations for SDEs with small diffusion and averaging, *Stochastic Process. Appl.*, **89** (2000), 69–79.

[29] Yin, G., Zhang, Q., *Continuous-time Markov Chians and Applications: A Two-Time-Scale Approach*, 2nd Ed., Springer, New York, NY, 2013.